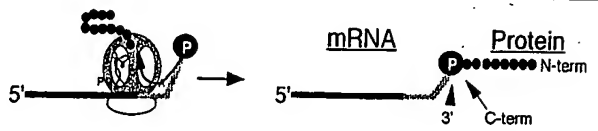


FIG. 1. Formation of mRNA–protein fusions on the ribosome. The ribosome pauses at an RNA/DNA junction, allowing puromycin to thread into the ribosome, entering the A site and forming fusion by accepting the nascent peptide from the peptidyl-tRNA in the P site.

united in a single molecule, RNA–protein fusions provide a means for reading and amplifying a protein sequence after it has been purified based on its function, effectively reverse translating the protein. The maximum library size that may be generated is limited by the size and efficiency of the translation reaction and by the efficiency of fusion formation on the ribosome. At the present time, libraries containing more than 10^{13} different sequences can be readily generated. Further, peptides and proteins synthesized as fusions commonly retain the binding properties of the unfused polypeptides. Finally, proteins ranging in size from 1 to at least 30 kDa can be synthesized as fusions, thus opening a great diversity of systems to examination.

Selection Scheme

The basic scheme in an RNA–protein fusion selection experiment is highlighted in Fig. 2, divided into 10 discrete steps: (1) generation of the

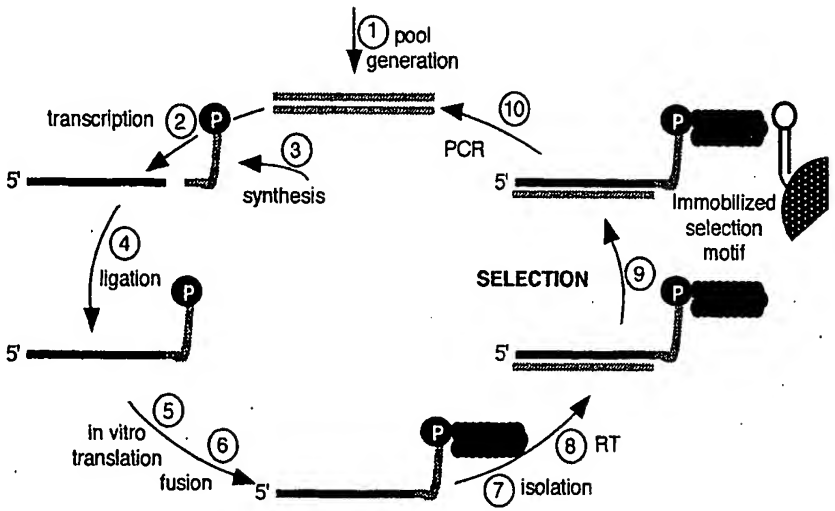


FIG. 2. Generalized selection scheme using mRNA–protein fusions. See text for description and optimization of individual steps.

initial double-stranded DNA sequence or pool, (2) transcription of the DNA into mRNA, (3) synthesis of the 3′-puromycin oligonucleotide, (4) ligation of the puromycin oligonucleotide to the mRNA, (5) *in vitro* translation of mRNA–puromycin templates, (6) generation of the mRNA–protein fusion, (7) isolation of the fusion, (8) reverse transcription to generate the cDNA/mRNA–protein fusion, (9) isolation of functional fusions with an immobilized selection motif, and (10) polymerase chain reaction (PCR) to generate an enriched dsDNA pool.

1. dsDNA Library

The starting library is constructed as a mixture of double-stranded DNA sequences. The DNA sequence contains several important design features. A T7 promoter is present at the 5′ end to allow large-scale synthesis of mRNA *in vitro* using T7 polymerase.[13] The transcript begins with 3 G nucleotides to aid transcription initiation. The remainder of the 5′-untranslated region (5′UTR) should be chosen according to the *in vitro* translation system to be used for fusion generation. For translation in reticulocyte lysate, we commonly use a deletion mutant of the tobacco mosaic virus 5′ UTR (ΔTMV) that provides efficient translation initiation.[12,14] For translation in *Escherichia coli,* a Shine–Dalgarno sequence appropriately spaced with respect to the start codon should be chosen.[15,16]

In contrast with bacteria, the 5′ UTR in eukaryotes does not contain a ribosome-binding site or an extensive 5′ consensus sequence. Although reports have been made of sequences that greatly enhance translation,[17] in reticulocyte lysate many sequences function efficiently as the 5′ UTR.[18] In general, eukaryotic translation systems use the first AUG codon in the mRNA to initiate protein synthesis. The precise sequence context surrounding this codon influences the efficiency of translation.[18,19] The sequence 5′RNNAUGR provides a good start context for most sequences, with a preference for A as the first purine (−3) and G as the second (+4).[18,20]

The open reading frame (ORF) can be constructed from either a defined sequence(s) or a random sequence library. The most important feature of the ORF and adjacent 3′ constant region is that neither contain stop codons.

[13] J. F. Milligan and O. C. Uhlenbeck, *Methods Enzymol.* **180,** 51 (1989).
[14] D. R. Gallie, D. E. Sleat, J. W. Watts, P. C. Turner, and T. M. A. Wilson, *Nucleic Acid Res.* **16,** 883 (1988).
[15] J. A. Steitz and K. Jakes, *Proc. Natl. Acad. Sci. U.S.A.* **72,** 4734 (1975).
[16] G. D. Stormo, T. D. Schneider, and L. M. Gold, *Nucleic Acids Res.* **10,** 2971 (1982).
[17] S. A. Jobling and L. Gehrke, *Nature* **325,** 622 (1987).
[18] M. Kozak, *Microbiol. Rev.* **47,** 1 (1983).
[19] M. Kozak, *J. Biol. Chem.* **266,** 19867 (1991).
[20] M. Kozak, *J. Mol. Biol.* **196,** 947 (1987).

Conclusions

In vitro selection of random rRNA fragments (SERF) is a simple and straightforward method to find the native rRNA-binding site for ribosomal proteins. As demonstrated, it has the potential to find a minimal rRNA fragment, which may be suitable for structural investigations by means of NMR spectroscopy or X-ray crystallography.

The atomic structure of 15 ribosomal proteins has been solved, and one of the surprising findings was that many of the proteins have two potential RNA-binding sites. It might be difficult to find both interactions in a single SERF experiment. However, in some cases the proposed binding sites are located on different domains, examples include the ribosomal proteins S4, S5, S8, L1, L6, and L9. One possible strategy is to separate the domains of those proteins by genetic means and to perform the selection with each domain separately.

Acknowledgments

We thank Dr. François Franceschi for kindly providing purified L11 from *Thermus thermophilus*, Sean Connell for help and discussions, and Detlev Kamp for expert assistance in the purification of *E. coli* ribosomal proteins.

[19] Optimized Synthesis of RNA–Protein Fusions for *in Vitro* Protein Selection

By RIHE LIU, JEFFREY E. BARRICK, JACK W. SZOSTAK, and RICHARD W. ROBERTS

Introduction

The extension of *in vitro* selection technology to the *in vitro* selection of peptides and proteins represents an area of great interest, in large part because the chemistry of molecular recognition and catalysis in living systems is largely mediated by polypeptides. The difficulty in designing schemes for *in vitro* protein selection is that while *in vitro* selection involves an iterated cycle of selection and amplification, there is no simple way to amplify protein molecules that have been selected for function. In order to isolate peptides or proteins with a desired function, the genetic information must be kept topologically linked to the protein in the form of a coding sequence such as RNA or DNA, a set of chemical tags, or a physical address. A major advantage of fully *in vitro* approaches is that they allow

0076-6879/00 $30.00

the isolation of proteins with desired properties even when no *in vivo* selection strategy exists or can be designed.

In vitro selection experiments begin with the generation of a population or pool containing many different sequences. This pool is then sieved to identify those individuals that have the desired functional properties. The number of different molecules that can be examined (the pool size or complexity) is one of the most important variables in a combinatorial experiment. For example, RNA molecules that bind ATP occur with a frequency of $1/10^{11}$ in random sequence RNA libraries.[1,2] Thus, an RNA selection designed to isolate ATP-binding aptamers would be unlikely to succeed if the starting library contained *only* 1 billion sequences.

Until recently, *in vivo* and *in vitro* protein selection experiments (e.g., the yeast two-hybrid system[3] and phage display[4–7]) were limited to complexities of about 1 million to 1 billion molecules, respectively. The main limitation on library size results from transfecting the starting cDNA library into the organism of choice. In contrast, *in vitro* RNA and DNA selection experiments routinely involve generating and screening libraries containing more than 10^{15} independent sequences. Two approaches have been developed that provide for totally *in vitro* selection of peptides and proteins: ribosome display and mRNA–protein fusions. Ribosome display, first examined in 1961,[8] has been the subject of previous reports and will not be covered in detail here.[9–11]

This article describes improvements in the development and use of mRNA–protein fusions for *in vitro* protein selection. An mRNA–protein fusion consists of a protein sequence covalently linked via its C terminus to the 3′ end of its own messenger RNA (Fig. 1).[12] The fusions are generated by *in vitro* translation of appropriate mRNA templates containing puromycin at their 3′ end. Because the coding and polypeptide sequences are

[1] M. Sassanfar and J. Szostak, *Nature* **364,** 550 (1993).
[2] D. H. Burke and L. Gold, *Nucleic Acids Res.* **25,** 2020 (1997).
[3] S. Fields and O.-K. Song, *Nature* **340,** 245 (1989).
[4] G. P. Smith, *Science* **228,** 1315 (1985).
[5] J. K. Scott and G. P. Smith, *Science* **249,** 386 (1990).
[6] J. J. Devlin, L. C. Panganiban, and P. E. Devlin, *Science* **249,** 404 (1990).
[7] S. E. Cwirla, E. A. Peters, R. W. Barrett, and W. J. Dower, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 6378 (1990).
[8] D. B. Cowie, S. Spiegelman, R. B. Roberts, and J. D. Duerksen, *Proc. Natl. Acad. Sci. U.S.A.* **47,** 114 (1961).
[9] L. C. Mattheakis, R. R. Bhatt, and W. J. Dower, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 9022 (1994).
[10] L. C. Mattheakis, J. M. Dias, and W. J. Dower, *Methods Enzymol.* **267,** 195 (1996).
[11] J. Hanes and A. Pluckthun, *Proc. Natl. Acad. Sci. U.S.A.* **94,** 4937 (1997).
[12] R. W. Roberts and J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **94,** 12297 (1997).

DUPLEX DOCUMENT INDEX SHEET

Serial Number

☐ 371P _____
PCT Papers in a 371P Application

☐ FOR _____
Foreign Reference

☐ NPL _____
Non-Patent Literature

☐ FRPR _____
Foreign Priority Papers

Date _____ Doc Code _____ Pages _____

Employee ID _____

DUPLEX

02/06/04 12:21 PM.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07B 61/00 | A2

(11) International Publication Number: **WO 00/21909**

(43) International Publication Date: 20 April 2000 (20.04.00)

(21) International Application Number: PCT/US99/23444

(22) International Filing Date: 7 October 1999 (07.10.99)

(30) Priority Data:
09/169,426 9 October 1998 (09.10.98) US

(71) Applicant: PHARMACOPEIA, INC. [US/US]; 3000 Eastpark Boulevard, Cranbury, NJ 08512 (US).

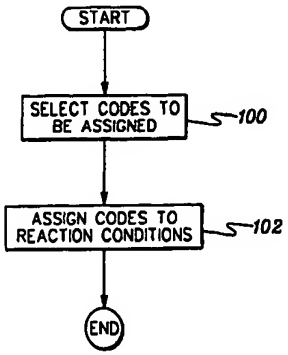
(72) Inventors: DILLARD, Lawrence, Wayne; 278 Springhill Road, Skillman, NJ 08558 (US). CONNELLY, James, Andrew; 1865 W. Desert Forest Court, Oro Valley, AZ 85737 (US). BALDWIN, John, J.; 621 Gypsy Hill Circle, Gwynedd Valley, PA 19437 (US). HORLBECK, Eric, George; 12784 Calle de la Siena, San Diego, CA 92130 (US). KIRK, Gregory, L.; 23 Jefferson Road, Winchester, MA 01890 (US). LAURI, Giorgio; 8 Arthurs Round Table, Wynnewood, PA 19096 (US).

(74) Agent: SCHILLER, Blanche, E.; Heslin & Rothenberg, P.C., 5 Columbia Circle, Albany, NY 12203 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SELECTING CODES TO BE USED FOR ENCODING COMBINATORIAL LIBRARIES



(57) Abstract

Codes to be used for encoding combinatorial libraries are selectively chosen based on one or more predefined function or criterion. In particular, a subset of N possible codes is selected based on some criterion. In one example, the codes are binary codes, and each code represents the tags used during a particular stage of synthesis of members of a combinatorial library. The tags define the reaction condition used during that particular stage of synthesis. In one embodiment, the predefined criterion ensures that each code includes more than one tag. This helps eliminate ambiguity during a decoding process in which the tags are identified to determine the reaction history during synthesis.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SELECTING CODES TO BE USED FOR ENCODING COMBINATORIAL LIBRARIES

TECHNICAL FIELD

This invention relates, in general, to the encoding of combinatorial libraries and, in particular, to selectively choosing codes to be assigned to reaction conditions used during synthesis of a combinatorial library.

BACKGROUND ART

Combinatorial techniques of chemical synthesis allow the creation of molecular libraries having immense diversity. These techniques entail a series of chemical steps with multiple choices of reaction conditions (e.g., reagents, temperature changes, etc.) for each step. The complexity, or number of members in a combinatorial library, is given by the product of the number of reaction conditions for each step of the synthesis, and can therefore, be quite large. The challenge in using combinatorial libraries is the characterization of members of the library with particular desired properties.

One solution to the above challenge is to use a split synthesis or direct divide technique to perform chemical synthesis on solid particles, such as beads. Through a protocol of separating and mixing beads during the synthesis, each bead in the final library has a product from a single, specific reaction sequence chemically bound to it, and that product is likely to differ from that bound to another bead.

During each step of the synthesis, zero or more tags (e.g., tagging molecules) are attached to each bead in order to encode the reaction condition used in that step, as well as the step number.

In one embodiment, the tags are used in combination with one another to form a binary record of the synthetic steps for each bead. For example, assume a combinatorial synthesis using any of seven different reagents in each of N steps is to be carried out. Such a combinatorial synthesis would yield 7^N different final products. As an example, the various reagents which can be used in any step are designated as binary 001(Reagent 1), 010(Reagent 2), 011(Reagent 3),... 111(Reagent 7). Thus, a binary synthesis code describing any complete N-STEP synthesis using 3 x N binary digits can be written.

For instance, if Reagent 3 is used in the first step, the binary numerical description is 011. If Reagent 1 is used in the second step, the description is 001 011. Further, if Reagent 6 is used in the third step, the description is 110 001 011. This 9-bit binary synthesis code describes the synthesis, and can be read from right to left in 3-bit blocks to decode the reagents used in each step of the synthesis.

To represent such a synthesis code chemically, a set of distinguishable, sensitively detectable molecules is used as tags, and the presence of a particular tag represents a binary "1" for the corresponding bit. Using a set of nine tagging molecules, T9-T1, for the above example, where T9 represents the leftmost binary bit and T1 represents the rightmost bit, the tag mixture containing only T9, T8, T4, T2 and T1 represents the 110 001 011 synthesis code.

The use of tags and various encoding techniques are described in detail in one or more of the following references, each of which is hereby incorporated herein by reference in its entirety: Ohlmeyer et al., "Complex synthetic chemical libraries indexed with molecular tags", Proceedings Of The National Academy Of Sciences Of The United States Of America, Vol. 90, No. 23, pp. 10922-10926 (December 1993); J.J. Baldwin, "Design, synthesis and use of binary encoded synthetic chemical libraries", Molecular Diversity, Vol. 2, No. 1/2, pp. 81-88 (October 1996); Burbaum et al., "A paradigm for drug discovery employing encoded combinatorial libraries", Proceedings Of The National Academy Of

Sciences Of The United States Of America, Vol. 92, No. 13, pp. 6027-6031 (June 1995); Still et al., U.S. Patent No. 5,565,324, entitled "Complex Combinatorial Chemical Libraries Encoded With Tags", issued on October 15, 1996; Baldwin et al., U.S. Patent No. 5,618,825, entitled "Combinatorial Sulfonamide Library", issued on April 08, 1997; Baldwin et al., U.S. Patent No. 5,663,046, entitled "Synthesis Of Combinatorial Libraries", issued on September 02, 1997; Still et al., International Publication No. WO 94/08051, entitled "Complex Combinatorial Chemical Libraries Encoded With Tags", International Publication Date April 14, 1994; Dower et al., U.S. Patent No. 5,639,603, entitled "Synthesizing And Screening Molecular Diversity", issued on June 17, 1997; and Dower et al., International Publication No. WO 93/06121, entitled "Method Of Synthesizing Diverse Collections Of Oligomers", International Publication Date April 01, 1993.

While binary coding has been established as a viable technique in encoding complex combinatorial libraries, there are some shortcomings with the present techniques, especially during decoding of the tags.

Decoding is performed in order to determine the reaction history of a particular bead. In particular, during decoding, any tag(s) attached to a bead is detached and identified to determine the particular conditions that occurred during synthesis. One technique for separating and identifying tags is known as Capillary Gas Chromatography (GC).

During decoding, it is sometimes difficult to determine whether a tag is present due to impurities of similar retention times or tags present in low amounts. Thus, the decoding becomes ambiguous, and it is difficult, under those circumstances, to determine the appropriate binary code that represents the reaction history.

Based on the foregoing, a need exists for a coding technique that eliminates ambiguous code reading. A further need exists for a capability that

enables the selective choosing of codes, from N possible codes, to be assigned to reaction conditions. A yet further need exists for a capability that guarantees the presence of enough tag peaks in a chromatogram that, even in the presence of significant variability in the absolute timing of the run, the relative timing can be determined. That is, a need exists for the presence of at least two tag peaks, so that the time distance (i.e., relative timing) between those peaks can be determined, even if there is significant variability in their positions (i.e., absolute timing). A further need exists for a capability that uses this time distance to confirm whether a particular peak represents the presence of a tag or a contaminant (e.g., substantially constant time spacing indicates the presence of a tag). The use of time spacing as a confirmation of whether a peak represents a tag or a contaminant is referred to herein as "self-clocking".

SUMMARY OF THE INVENTION

The shortcomings of the prior art are overcome and additional advantages are provided through the provision of a method of determining codes usable in encoding combinatorial libraries. The method includes, for instance, selecting a plurality of codes to be assigned to a plurality of reaction conditions usable during synthesis of a combinatorial library. Each of the plurality of codes includes a plurality of tags, wherein none of the plurality of codes includes only a single tag. The method further includes assigning selected codes to reaction conditions.

In one embodiment, each of the plurality of codes is a binary code, and a binary "one" within the binary code represents the presence of a particular tag.

In another embodiment of the invention, the plurality of codes is selected using a predefined criterion. The predefined criterion specifies at least one of the following: each of the plurality of codes includes an even number of tags present therein; each of the plurality of codes includes an odd number of tags present therein; each of the plurality of codes includes up to a maximal number of tags

present therein; each of the plurality of codes includes up to a maximal number of "zero" bits; and each of the plurality of codes does not include a predetermined pattern of bits.

In another embodiment, the plurality of codes is selected using a parity bit.

In a further aspect of the present invention, a method of determining codes usable in encoding chemical libraries is provided. The method includes selecting, from N possible codes, a group of codes to be assigned to a plurality of reaction conditions usable during synthesis of a chemical library. The selecting includes using a predefined function to select the group of codes, wherein the predefined function selects fewer than N-1 codes from the N possible codes. The method further includes assigning selected codes to reaction conditions.

In yet a further aspect of the present invention, a method of determining codes usable in encoding chemical libraries is provided. The method includes selecting, from N possible codes, a plurality of codes to be assigned to a plurality of reaction conditions, wherein the plurality of codes satisfies a predefined criterion. The predefined criterion is other than excluding an "all zeroes" code. The method further includes assigning selected codes to reaction conditions.

In a further aspect of the present invention, a system of determining codes usable in encoding combinatorial libraries is provided. The system includes means for selecting a plurality of codes to be assigned to a plurality of reaction conditions, in which each of the plurality of codes includes a plurality of tags, such that none of the plurality of codes includes only a single tag. The system further includes means for assigning selected codes to reaction conditions.

In another aspect of the present invention, a system of determining codes usable in encoding chemical libraries is provided. The system includes means for selecting, from N possible codes, a group of codes to be assigned to a plurality of

reaction conditions usable during synthesis of a chemical library. The means for selecting includes means for using a predefined function to select the group of codes, wherein the predefined function selects fewer than N-1 codes from the N possible codes. The system further includes means for assigning selected codes to reaction conditions.

In yet a further aspect of the present invention, a system of determining codes usable in encoding chemical libraries is provided. The system includes means for selecting, from N possible codes, a plurality of codes to be assigned to a plurality of reaction conditions, wherein the plurality of codes satisfies a predefined criterion, wherein the predefined criterion is other than excluding an "all zeroes" code. The system further includes means for assigning selected codes to reaction conditions.

In another aspect of the present invention, an article of manufacture is provided, including at least one computer usable medium having computer readable program code means embodied therein for causing the determining of codes usable in encoding combinatorial libraries. The computer readable program code means includes computer readable program code means for causing a computer to select a plurality of codes to be assigned to a plurality of reaction conditions usable during synthesis of a combinatorial library, each of the plurality of codes including a plurality of tags, wherein none of the plurality of codes includes only a single tag; and computer readable program code means for causing a computer to assign selected codes to reaction conditions.

In yet another aspect of the present invention, at least one program storage device is provided, which is readable by a machine, tangibly embodying at least one program of instructions executable by the machine to perform a method of determining codes usable in encoding chemical libraries. The method includes selecting, from N possible codes, a group of codes to be assigned to a plurality of reaction conditions usable during synthesis of a chemical library. The selecting

includes using a predefined function to select the group of codes, wherein the predefined function selects fewer than N-1 codes from the N possible codes. The method further includes assigning selected codes to reaction conditions.

In a further aspect of the present invention, at least one program storage device is provided, which is readable by a machine, tangibly embodying at least one program of instructions executable by the machine to perform a method of determining codes usable in encoding chemical libraries. The method includes selecting, from N possible codes, a plurality of codes to be assigned to a plurality of reaction conditions, wherein the plurality of codes satisfies a predefined criterion. The predefined criterion is other than excluding an "all zeroes" code. The method further includes assigning selected codes to reaction conditions.

In accordance with the principles of the present invention, a coding capability is provided that eliminates ambiguous code reading. Further, the capability of the present invention advantageously enables the selective choosing of codes, from N possible codes, to be assigned to reaction conditions. Further, the capability of the present invention guarantees the presence of enough tag peaks in a chromatogram that, even in the presence of significant variability in the absolute timing of the run, the relative timing can be determined. Additionally, the present invention is advantageously self-clocking. The use of the present invention in each synthesis step can result in the avoidance of single bit codes for reaction conditions.

Additional features and advantages are realized through the techniques of the present invention. Other embodiments and aspects of the invention are described in detail herein and are considered a part of the claimed invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The subject matter which is regarded as the invention is particularly pointed out and distinctly claimed in the claims at the conclusion of the specification. The foregoing and other objects, features, and advantages of the invention will be apparent from the following detailed description taken in conjunction with the accompanying drawings in which:

FIG. 1 depicts a generalized method of the present invention;

FIG. 2 depicts one embodiment of the logic associated with the selection capability of the present invention;

FIG. 3 depicts one example of N possible codes, in which a subset is selected (or "accepted") therefrom, in accordance with the principles of the present invention;

FIG. 4 depicts one example of a table of accepted codes, in accordance with the principles of the present invention;

FIG. 5 depicts one embodiment of the table of accepted codes of FIG. 4, in which the codes are assigned to reaction conditions, in accordance with the principles of the present invention;

FIG. 6 depicts another example of accepted codes, in accordance with the principles of the present invention;

FIG. 7a depicts another example of N possible codes, in which a subset is selected (or accepted) therefrom, in accordance with the principles of the present invention;

FIG. 7b depicts one example of adding a parity bit to the N possible codes of FIG. 7a in order to make a selection of the codes to be

included in the subset of accepted codes, in accordance with the principles of the present invention;

FIG. 8 depicts one example in which it is difficult to determine whether a tag is present in a sample;

FIG. 9 illustrates one example of output from a gas chromatograph, in accordance with the principles of the present invention;

FIG. 10 depicts another example of output from a gas chromatograph, in accordance with the principles of the present invention; and

FIG. 11 depicts one embodiment of a computer environment providing and/or using the capability of the present invention.

BEST MODE FOR CARRYING OUT THE INVENTION

In accordance with the principles of the present invention, a capability is provided in which codes to be assigned to reaction conditions are selectively chosen based on one or more criterion. That is, out of N possible codes, a subset of the N codes is selected based on some constraint or some predefined function.

The codes are, for instance, binary codes, and each code represents the tags used during a particular stage of synthesis of members of a combinatorial library. Specifically, the tags define the reaction condition used during that particular stage of synthesis.

A generalized technique of one embodiment of the present invention is described with reference to FIG. 1. Initially, a group of acceptable codes is selectively chosen, from N possible codes, based upon a predefined function or

one or more criterion, as described in detail below, STEP 100. Thereafter, the selected codes are assigned to reaction conditions, and those conditions may be used during synthesis of library members, STEP 102. The assignment may be arbitrary or may be based on one or more factors. For instance, the first code may be assigned to the first condition to be used during synthesis, etc.

One embodiment of the selection process used to choose a group of codes to be assigned to the reaction conditions is described in detail with reference to FIG. 2. Initially, a decision is made as to the N possible codes that could be used for a particular synthesis step, STEP 200. For example, a determination is made as to how many bits are sufficient to represent the number of reaction conditions to be used during that step. Assume for this one example that four bits are sufficient. Thus, in a binary scheme, there are 2^4 possible codes (i.e., N=16), as shown in FIG. 3.

Out of the possible codes, only those codes that satisfy one or more criterion (or predefined function) are selected. The criterion can include many different possibilities. As examples, the criterion can select codes having an even number of "one" bits (i.e., an even number of tags); an odd number of "one" bits; an odd number, greater than one, of "one" bits; more than one tag (i.e., more than one binary "one" bit); up to a total number of "zero" bits; up to a total number of "one" bits; up to a maximal allowed sequence of "zero" or "one" bits; an even parity; an odd parity; codes that do not include a predefined sequence of bits (e.g., the codes that do not include a sequence of 101), etc. The above criteria are only provided as examples. There are many more possibilities, and each of those possibilities is considered a part of the claimed invention.

For illustration purposes only, the predefined function, in this example, includes all those codes that have up to a total of two "zeroes". Thus, the process continues with determining which of the codes of FIG. 3 meet this criterion.

Returning to FIG. 2, a code is considered from the possible codes, STEP 202. For example, code 0000 is considered from the codes depicted in FIG. 3. A determination is made as to whether the considered code meets the criterion of having no more than two zeroes, STEP 204. Since code 0000 has more than two zeroes, it does not meet the criterion, and thus, is not an acceptable code. Therefore, if there are further codes to be considered, INQUIRY 206, a new code is considered, STEP 202.

Proceeding sequentially down the list of the possible codes illustrated in FIG. 3, the next code is 0001. Again, since code 0001 does not meet the criterion, STEP 204, and there are more codes, INQUIRY 206, another code is selected. This procedure continues. At some point, a code is selected that does meet the criterion, INQUIRY 204. For instance, code 0011 meets the criterion of having at a maximum two "zeroes", thus, this code is selected as an acceptable code.

In one example, the acceptable code is placed in a table of acceptable codes, STEP 208. One instance of such a table is depicted in FIG. 4. The above process continues until all of the acceptable codes are selected from the possible codes based on the predefined criterion.

As described with reference to FIG. 1, after the acceptable codes are determined, each of the acceptable codes can be assigned, respectively, to one of a plurality of reaction conditions, as depicted in FIG. 5. In this particular case, ten reaction conditions are represented by ten unique codes.

Another example of a predefined function used to select acceptable codes from N possible codes is one in which all codes having an even number of tags is selected. This is depicted in FIG. 6. In this particular example, five bits are considered necessary to represent the various reaction conditions to be used during synthesis. Thus, there are 32 possible codes. Out of the 32 possible codes,

15 codes are acceptable (designated by "used" in the table). That is, 15 codes meet the criterion of having an even number of tags.

In yet another embodiment of the present invention, parity bits may be added to the N possible codes in order to determine the acceptable codes. For example, assume there are eight possible codes, as shown in FIG. 7a. Further, assume that even parity is to be used, in this example. Thus, for each code in which the addition of a binary "one" provides an even number of "one" bits, a parity "one" bit is added, as shown in FIG. 7b.

Thereafter, the codes having the parity "one" bit are selected from the N possible codes. In the example depicted in FIG. 7b, the following codes are chosen: 0011, 0101, 1001 and 1111.

In the above examples in which each code is a binary code, each binary "one" represents the existence of a tag and each binary "zero" represents the absence of a tag. When a reaction condition is used in a particular synthesis step, the tags represented by the binary code associated with that reaction condition are also added during that synthesis step. This provides a record of the conditions used during synthesis of a particular library member.

For example, assume Reaction Condition 3 (FIG. 5) is used during a first synthesis step, Reaction Condition 1 is used during a second step, and Reaction Condition 6 is used during a third step, then the synthesis code representing the three reaction conditions is 0111 0011 0110. This synthesis code can be read from right to left, in 4-bit blocks, to decode the reaction conditions used during each step of the synthesis. (In another embodiment, the 10 different reaction conditions in FIG. 5 would be alternatives for a single reaction step. A different set of reaction conditions, encoded by a different set of tags, would be used for the next reaction step.)

To represent the synthesis code chemically, a set of distinguishable tags is used, in which the presence of a particular tag is represented by a binary "1". In this specific example, twelve (12) bits were used to represent 3 reaction conditions and 3 steps and, thus, a set of twelve tagging molecules are used, T12-T1. T12 represents the leftmost binary bit and T1 represents the rightmost bit of the synthesis code. Therefore, the following tags would represent the 0111 0011 0110 synthesis code: T11, T10, T9, T6, T5, T3 and T2.

The manner in which tagging molecules are prepared and the manner in which tagging molecules relate to the binary bits of a synthesis code, are known in the art and described in various of the above references, each of which has been incorporated herein by reference in its entirety.

When the reaction history of a particular bead is desired, the tags associated with that bead are decoded. In one example, a chromatogram is produced showing peaks where tags are present. From the peaks, a binary code is determined. For instance, in the above example, since a chromatogram would show that T2 and T3 are present and T1 and T4 are absent, the binary code 0110 is provided, This code indicates that Reaction Condition 3 was used during the first synthesis step.

Sometimes the peaks of the chromatogram are not well defined, and thus, it is difficult to determine whether a tag is present. For instance, it is possible that during the tagging reaction, one tag may not be incorporated in a particular bead at the same quantitative level as others within the set. It is also possible that during the detagging/gas chromatograph (GC) portion of the process, some impurities can be introduced into the tag mixture, which may have retention times similar to that of one of the tags in the set. Both of these occurrences can result in chromatogram data in which the tagging code is ambiguous. An example of such is depicted in FIG. 8.

In the example of FIG. 8, a 5 place binary code is represented. Clearly, tags 5 and 2 are present, while tags 1 and 3 are not. However, because the peak having the retention time in the expected range for tag 4 is of much lower height than either of the other two, its identity is in question. As a result, the binary code, from left to right, can be either 11010 or 10010.

To eliminate this type of ambiguous code reading, the encoding protocol of the present invention is used, in which a specially chosen subset of the N possible binary codes is employed. For example, if a group of codes is selected from the 32 (2^5) (see FIG. 6) possible codes, based on a selection criterion of only those codes where the sum of the individual digits is even, then 11010 would not be an acceptable code. Thus, 10010 must be the code that represents the sample depicted in FIG. 8. With this strategy, $2^{(N-1)}$-1 unique sets of conditions can be encoded with N tags.

Further examples in which the selective code capability of the present invention is used are depicted in FIGs. 9-10. Each of the figures depicts a sample chromatogram, and each tag within the chromatogram is labelled by CnnCLn. For example, in FIG. 9, peak 900 has tag C12CL5.

In FIG. 9, there is an ambiguity as to whether peak 902 represents the presence of a particular tag or not. Thus, the binary code for that step may be 1101 or 1001. However, based on the criterion for acceptable codes used in this example (e.g., an even number of tags), binary code 1001 (reference number 904) must be the correct code. Therefore, in accordance with the present invention, that peak does not have a tag.

Similarly, in FIG. 10, peak 1000 is in question. However, based on the present invention in which the criterion specifies no single tags, the peak is determined to be tagged. Thus, a binary "one" represents that peak (see reference number 1002). The same holds true for peak 1004.

The capability of the present invention can readily be automated by creating a suitable program, in software, hardware, microcode, firmware or any combination thereof. Further, any type of computer or computer environment can be employed to provide, incorporate and/or use the capability of the present invention. One such environment is depicted in FIG. 11 and described in detail below.

In one embodiment, a computer environment 1100 includes, for instance, at least one central processing unit 1102, a main storage 1104, and one or more input/output devices 1106, each of which is described below.

As is known, central processing unit 1102 is the controlling center of computer environment 1100 and provides the sequencing and processing facilities for instruction execution, interruption action, timing functions, initial program loading and other machine related functions. The central processing unit executes at least one operating system, which as known, is used to control the operation of the computing unit by controlling the execution of other programs, controlling communication with peripheral devices and controlling use of the computer resources.

Central processing unit 1102 is coupled to main storage 1104, which is directly addressable and provides for high speed processing of data by the central processing unit. Main storage may be either physically integrated with the CPU or constructed in stand alone units.

Main storage 1104 is also coupled to one or more input/output devices 1106. These devices include, for instance, keyboards, communications controllers, teleprocessing devices, printers, magnetic storage media (e.g., tape, disks), direct access storage devices, and sensor based equipment. Data is transferred from main storage 1104 to input/output devices 1106, and from the input/output devices back to main storage.

Described in detail above is an improvement of the coding process of combinatorial libraries, in which group coding is used. Group coding includes selectively choosing a subset of codes, from N possible codes, that meets one or more predetermined criterion. The subset can include codes that are selected based on a parity bit or by some other mechanism.

In one example, the code selection capability of the present invention guarantees the presence of enough tag peaks in the chromatogram that, even in the presence of significant variability in the absolute timing of the run, the relative timing can be determined and the "zero" bits identified reliably. The group coding thus is considered self-clocking.

In one embodiment, the selection capability of the present invention allows more than $2^{(N-1)}$-1 possibilities for N bits. Further, no bit can be claimed to be merely extraneous to the code. A bit is considered extraneous to the code when the bit is added to the code just for the sake of adding a bit and that bit can really be ignored when the code is read. With the present invention, these extraneous bits are avoided and thus, more efficient use of tags can be made.

The use of the present invention at each synthesis step can result in avoiding single-bit codes for reaction conditions. Additionally, it advantageously allows for more reliable interpretation of an individual chromatogram by using the guaranteed presence of a minimum number of tags to create an "internal standard" for the shifts in that chromatogram. Further, it allows for independent error checking of the validity of the tags at each step based on the frequency of the code occurrence and the identification of invalid codes.

Although the examples described above reference binary coding, the present invention is also applicable to higher order coding or other types of coding. Thus, these are considered a part of the claimed invention.

The present invention can be included in an article of manufacture (e.g., one or more computer program products) having, for instance, computer usable media. The media has embodied therein, for instance, computer readable program code means for providing and facilitating the capabilities of the present invention. The article of manufacture can be included as a part of a computer system or sold separately.

Additionally, at least one program storage device readable by a machine, tangibly embodying at least one program of instructions executable by the machine to perform the capabilities of the present invention can be provided.

The flow diagrams depicted herein are just exemplary. There may be many variations to these diagrams or the steps (or operations) described therein without departing from the spirit of the invention. For instance, the steps may be performed in a differing order, or steps may be added, deleted or modified. All of these variations are considered a part of the claimed invention.

Although preferred embodiments have been depicted and described in detail herein, it will be apparent to those skilled in the relevant art that various modifications, additions, substitutions and the like can be made without departing from the spirit of the invention and these are therefore considered to be within the scope of the invention as defined in the following claims.

CLAIMS

What is claimed is:

1. A method of determining codes usable in encoding combinatorial libraries, said method comprising:

selecting a plurality of codes to be assigned to a plurality of reaction conditions usable during synthesis of a combinatorial library, each of said plurality of codes comprising a plurality of tags, wherein none of said plurality of codes comprises only a single tag; and

assigning selected codes to reaction conditions.

2. The method of claim 1, wherein each of said plurality of codes is a binary code, and wherein a binary "one" within said binary code represents the presence of a particular tag.

3. The method of claim 1, wherein said selecting comprises using a predefined criterion to select said plurality of codes from N possible codes.

4. The method of claim 3, wherein said predefined criterion specifies at least one of the following:

each of said plurality of codes includes an even number of tags present therein;

each of said plurality of codes includes an odd number of tags present therein, wherein said odd number is greater than one;

each of said plurality of codes includes up to a maximal number of tags present therein;

each of said plurality of codes includes up to a maximal number of "zero" bits; and

each of said plurality of codes does not include a predetermined pattern of bits.

5. The method of claim 1, wherein said selecting comprises using a parity bit to determine which codes of N possible codes are to be selected as said plurality of codes.

6. A method of determining codes usable in encoding chemical libraries, said method comprising:

selecting, from N possible codes, a group of codes to be assigned to a plurality of reaction conditions usable during synthesis of a chemical library, said selecting comprising using a predefined function to select said group of codes, wherein said predefined function selects fewer than N-1 codes from said N possible codes; and

assigning selected codes to reaction conditions.

7. The method of claim 6, wherein said predefined function specifies that each code of said group of codes has up to a maximal number of "zero" bits.

8. The method of claim 6, wherein said predefined function specifies that each code of said group of codes has up to a maximal number of tags.

9. The method of claim 6, wherein said predefined function specifies that each code of said group of codes has up to a maximal number of "one" bits.

10. The method of claim 6, wherein said predefined function specifies that each code of said group of codes has an even number of tags present.

11. The method of claim 6, wherein said predefined function specifies that each code of said group of codes has an odd number of tags present.

12. The method of claim 11, wherein said predefined function specifies that each code of said group of codes has an odd number, greater than one, of tags present.

13. The method of claim 6, wherein said predefined function comprises not including in said group of codes any code having a predetermined pattern.

14. The method of claim 6, wherein said predefined function comprises using a parity bit to select said group of codes.

15. The method of claim 6, wherein each of said plurality of codes is a binary code, and wherein a binary "one" within said binary code represents the presence of a particular tag.

16. A method of determining codes usable in encoding chemical libraries, said method comprising:

selecting, from N possible codes, a plurality of codes to be assigned to a plurality of reaction conditions usable during synthesis of a chemical library, wherein said plurality of codes satisfies a predefined criterion, said predefined criterion being other than excluding an "all zeroes" code; and

assigning selected codes to reaction conditions.

17. The method of claim 16, wherein each of said plurality of codes is a binary code, and wherein a binary "one" within said binary code represents the presence of a particular tag.

18. The method of claim 16, wherein said predefined criterion specifies at least one of the following:

each of said plurality of codes includes an even number of tags present therein;

each of said plurality of codes includes an odd number of tags present therein;

each of said plurality of codes includes up to a maximal number of tags present therein;

each of said plurality of codes includes up to a maximal number of "zero" bits; and

each of said plurality of codes does not include a predetermined pattern of bits.

19. The method of claim 16, wherein said selecting comprises using a parity bit to determine which codes of N possible codes are to be selected as said plurality of codes.

20. A system of determining codes usable in encoding combinatorial libraries, said system comprising:

means for selecting a plurality of codes to be assigned to a plurality of reaction conditions usable during synthesis of a combinatorial library, each of said plurality of codes comprising a plurality of tags, wherein none of said plurality of codes comprises only a single tag; and

means for assigning selected codes to reaction conditions.

21. The system of claim 20, wherein said means for selecting comprises means for using a predefined criterion to select said plurality of codes from N possible codes.

22. A system of determining codes usable in encoding chemical libraries, said system comprising:

means for selecting, from N possible codes, a group of codes to be assigned to a plurality of reaction conditions usable during synthesis of a chemical library, said means for selecting comprising means for using a predefined function to select said group of codes, wherein said predefined function selects fewer than N-1 codes from said N possible codes; and

means for assigning selected codes to reaction conditions.

23. The system of claim 22, wherein said predefined function specifies that each code of said group of codes has up to a maximal number of "zero" bits.

24. The system of claim 22, wherein said predefined function specifies that each code of said group of codes has up to a maximal number of tags.

25. The system of claim 22, wherein said predefined function specifies that each code of said group of codes has an even number of tags present.

26. The system of claim 22, wherein said predefined function specifies that each code of said group of codes has an odd number of tags present.

27. The system of claim 22, wherein said predefined function comprises not including in said group of codes any code having a predetermined pattern.

28. The system of claim 22, wherein said predefined function comprises using a parity bit to select said group of codes.

29. A system of determining codes usable in encoding chemical libraries, said system comprising:

means for selecting, from N possible codes, a plurality of codes to be assigned to a plurality of reaction conditions usable during synthesis of a chemical library, wherein said plurality of codes satisfies a predefined criterion, said predefined criterion being other than excluding an "all zeroes" code; and

means for assigning selected codes to reaction conditions.

30. The system of claim 29, wherein said predefined criterion specifies at least one of the following:

each of said plurality of codes includes an even number of tags present therein;

each of said plurality of codes includes an odd number of tags present therein;

each of said plurality of codes includes up to a maximal number of tags present therein;

each of said plurality of codes includes up to a maximal number of "zero" bits; and

each of said plurality of codes does not include a predetermined pattern of bits.

31. The system of claim 29, wherein said means for selecting comprises means for using a parity bit to determine which codes of N possible codes are to be selected as said plurality of codes.

32. An article of manufacture, comprising:

at least one computer usable medium having computer readable program code means embodied therein for causing the determining of codes usable in encoding combinatorial libraries, the computer readable program code means in said article of manufacture comprising:

computer readable program code means for causing a computer to select a plurality of codes to be assigned to a plurality of reaction conditions usable during synthesis of a combinatorial library, each of said plurality of codes comprising a plurality of tags, wherein none of said plurality of codes comprises only a single tag; and

computer readable program code means for causing a computer to assign selected codes to reaction conditions.

33. At least one program storage device readable by a machine, tangibly embodying at least one program of instructions executable by the machine to perform a method of determining codes usable in encoding chemical libraries, said method comprising:

selecting, from N possible codes, a group of codes to be assigned to a plurality of reaction conditions usable during synthesis of a chemical library, said selecting comprising using a predefined function to select said group of codes, wherein said predefined function selects fewer than N-1 codes from said N possible codes; and

assigning selected codes to reaction conditions.

34. At least one program storage device readable by a machine, tangibly embodying at least one program of instructions executable by the machine to perform a method of determining codes usable in encoding chemical libraries, said method comprising:

selecting, from N possible codes, a plurality of codes to be assigned to a plurality of reaction conditions usable during synthesis of a chemical library, wherein said plurality of codes satisfies a predefined criterion, said predefined criterion being other than excluding an "all zeroes" code; and

assigning selected codes to reaction conditions.

* * * * *

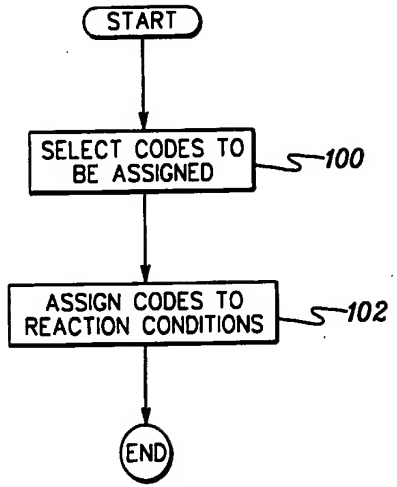


fig. 1

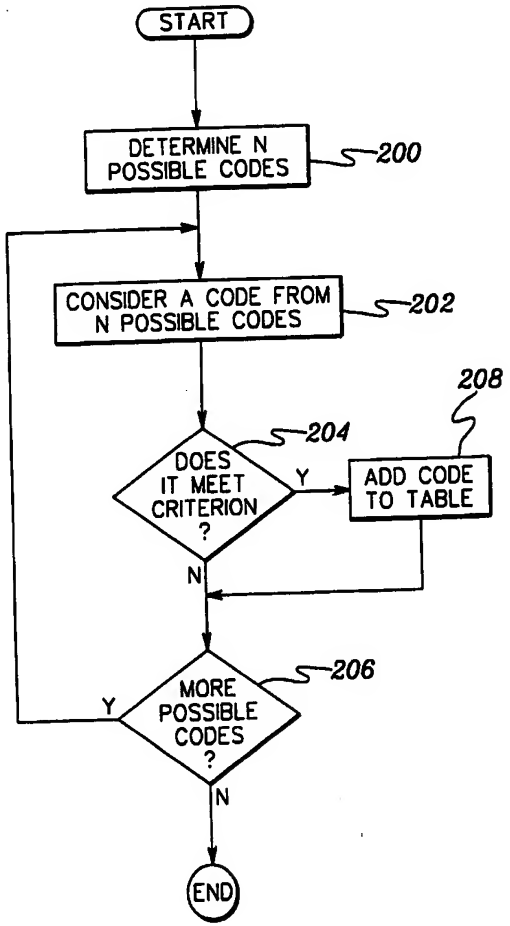


fig. 2

N POSSIBLE CODES

0000	1000
0001	1001
0010	1010
0011	1011
0100	1100
0101	1101
0110	1110
0111	1111

fig. 3

TABLE OF ACCEPTED CODES

CODE	REACTION CONDITION
0011	
0101	
0110	
1001	
1100	
0111	
1011	
1101	
1110	
1111	

fig. 4

TABLE OF ACCEPTED CODES

CODE	REACTION CONDITION
0011	CONDITION 1
0101	CONDITION 2
0110	CONDITION 3
1001	CONDITION 4
1100	CONDITION 5
0111	CONDITION 6
1011	CONDITION 7
1101	CONDITION 8
1110	CONDITION 9
1111	CONDITION 10

fig. 5

0	0	0	0	0	NOT USED
0	0	0	0	1	NOT USED
0	0	0	1	0	NOT USED
0	0	0	1	1	USED
0	0	1	0	0	NOT USED
0	0	1	0	1	USED
0	0	1	1	0	USED
0	0	1	1	1	NOT USED
0	1	0	0	0	NOT USED
0	1	0	0	1	USED
0	1	0	1	0	USED
0	1	0	1	1	NOT USED
0	1	1	0	0	USED
0	1	1	0	1	NOT USED
0	1	1	1	0	NOT USED
0	1	1	1	1	USED
1	0	0	0	0	NOT USED
1	0	0	0	1	USED
1	0	0	1	0	USED
1	0	0	1	1	NOT USED
1	0	1	0	0	USED
1	0	1	0	1	NOT USED
1	0	1	1	0	NOT USED
1	0	1	1	1	USED
1	1	0	0	0	USED
1	1	0	0	1	NOT USED
1	1	0	1	0	NOT USED
1	1	0	1	1	USED
1	1	1	0	0	NOT USED
1	1	1	0	1	USED
1	1	1	1	0	USED
1	1	1	1	1	NOT USED

fig. 6

POSSIBLE CODES
000
001
010
011
100
101
110
111

fig. 7a

CODE	PARITY
000	0
001	1
010	1
011	0
100	1
101	0
110	0
111	1

fig. 7b

fig. 8

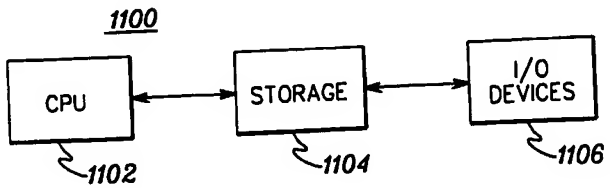


fig. 11

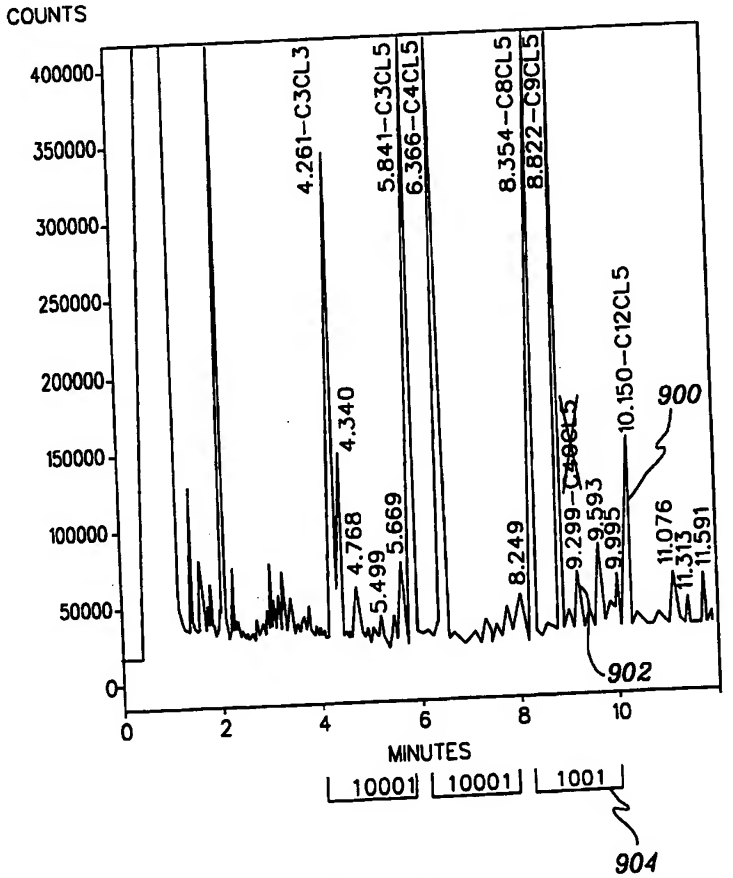


fig. 9